

基于近红外透射光谱及多种数据降维方法的红富士苹果产地溯源

马永杰¹, 郭俊先¹, 郭志明², 黄华³, 史勇¹, 周军¹

(1. 新疆农业大学机电工程学院, 新疆乌鲁木齐 830052) (2. 江苏大学食品与生物工程学院, 江苏镇江 212013)
(3. 新疆农业大学数理学院, 新疆乌鲁木齐 830052)

摘要: 为了探寻苹果产地溯源新方法, 本文以新疆阿克苏、陕西洛川、山东烟台三个不同产地 600 个红富士苹果样本为试材分别采集其 590~1250 nm 的近红外透射光谱图, 将经过光谱校正后的光谱数据做归一化 (Normalization)、多元散射校正 (Multivariate Scattering Correction, MSC) 等 8 种光谱预处理, 对经过预处理后的光谱数据建立全波长分类模型发现二阶求导为最优预处理方法; 然后再将经过二阶求导预处理的光谱数据结合以欧几里得距离 (Euclidean)、相关距离 (correlation)、余弦相似度 (cosine)、城市街区距离 (cityblock) 作为距离度量的 K 最近邻法建模, 发现 K 最近邻法 (correlation) 为最优分类识别方法; 随后再分别用高斯过程隐变量模型 (Gaussian Process Latent Variable Model, GPLVM)、线性局部切空间排列 (Linear Local Tangent Space Alignment, LLTSA)、等 12 种数据降维方法对经二阶求导预处理后的光谱做降维处理, 并结合 K 最近邻法 (correlation) 鉴别苹果产地。结果表明, 提取前 9 个主成分, 采用二阶求导-扩散映射-K 最近邻法 (correlation) 模型识别效果最优, 建模集和预测集的分类识别率分别为 97.30% 与 92.30%。故, 深度学习数据降维方法结合近红外透射光谱技术可成功、有效地实现苹果产地溯源。

关键词: 数据降维; 近红外光谱; 苹果; 产地溯源

文章编号: 1673-9078(2020)06-303-309

DOI: 10.13982/j.mfst.1673-9078.2020.6.1249

Origin Tracing of Red Fuji Apple Based on Near Infrared Transmission Spectrum and Various Dimension Reduction Methods

MA Yong-jie¹, GUO Jun-xian¹, GUO Zhi-ming², HUANG Hua³, SHI Yong¹, ZHOU Jun¹

(1.College of Mechanical and Electronic Engineering, Xinjiang Agricultural University, Urumqi 830052, China)

(2.College of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

(3.College of Mathematics and Physics, Xinjiang Agricultural University, Urumqi 830052, China)

Abstract: In order to find a new method of tracing apple origin, in this work, 600 apple samples from three different producing areas of Aksu, Luochuan, and Yantai were used to collect the near-infrared transmission spectra within 590~1250 nm, then spectroscopically corrected spectrum were treated by eight species of spectral pretreatment such as normalization, standard normal variate transformation, multivariate scattering correction, savitzky-golay smoothing, 2nd derivative, mean centering, moving average, 1st derivative. Firstly, established full spectra classification model for pre-processed spectral data showed that 2nd derivative was the best pre-processing method. Secondly, data set preprocessed by 2nd derivative were used to combine 4 different KNN models (Euclidean, correlation, cosine, cityblock) to do pattern recognition, which was found that K-nearest neighbor method (correlation) was the best classification and recognition method. Thirdly, factor analysis, gaussian process latent variable model, linear local tangent space alignment, neighborhood components analysis, neighborhood

引文格式:

马永杰,郭俊先,郭志明,等.基于近红外透射光谱及多种数据降维方法的红富士苹果产地溯源[J].现代食品科技,2020,36(6):303-309

MA Yong-jie, GUO Jun-xian, GUO Zhi-ming, et al. Origin tracing of red Fuji apple based on near infrared transmission spectrum and various dimension reduction methods [J]. Modern Food Science and Technology, 2020, 36(6): 303-309

收稿日期: 2019-12-18

基金项目: 国家自然科学基金项目 (61367001)

作者简介: 马永杰 (1995-), 男, 硕士研究生, 研究方向: 农产品无损快速检测

通讯作者: 郭俊先 (1975-), 男, 博士, 教授, 研究方向: 农产品无损快速检测

preserving embedding, diffusion maps, t-distributed stochastic neighbor embedding, landmark isomap, laplacian eigenmaps, locally linear embedding, principal component analysis, linear discriminant analysis were used to reduce the dimension of the spectrum after 2nd derivative pretreatment, and then combining K-nearest neighbor was combined to trace the origin of apple. Results showed that an optimal identification model was obtained by using 2nd derivative-diffusion maps-KNN (correlation). The identification rates for the calibration set and prediction set were 97.3% and 92.3%, respectively. Therefore, the deep learning dimension reduction methods combined with near-infrared transmission spectroscopy could successfully and effectively discriminate the traceability of apple origin.

Key words: dimension reduction; near infrared spectroscopy; apple; origin traceability

苹果属蔷薇科落叶乔木,全世界有一千种以上的栽培品种。我国现有栽植的主要品种有从欧洲、美洲、日本传输过来,也有自己培育的新品种。现在我国苹果的生产以乾县红富士、辽南寒富、山西万荣、甘肃天水花牛苹果、山东红星、陕西洛川富士以及新疆阿克苏红富士等为重点苹果产区^[1]。2017至2018年度我国苹果产量占世界苹果总产量的58.39%,是世界上最大的苹果栽培国家,也是产量最多的国家^[2]。

海内外众多专家及研究人员对苹果的成分及含量做了大量研究报告,研究结果表明同种苹果在不同产地的主要成分含量上的确存在一定的差异^[3]。这预示着苹果的化学组成部分的含量不仅与苹果的品种有关,也同样取决于苹果的产地。中国作为苹果生产大国,有众多获得原产地保护的国家地理标志产品,这是由于不同产地的苹果受到不同气候、环境等众多外界因素影响,其内部组成的主要成分含量存在差异,造成苹果在口感、水分、糖分等方面出现较为明显的差别,使其各自形成各自独特的风味,市场价格也随之不同。而正是由于市场对这些名优产品的大量需求,随之而来的就是各种鱼目混珠、乱掺乱卖的现象屡见不鲜,而通过颜色、气味、外形等传统鉴别方法观察已经无法准确判断苹果产地,导致市场混乱、对国家地理标志产品的声誉和消费者权益都造成严重的侵害。故,迫切需求一种简单、快捷、无损的方法对苹果产地准确检验。

近些年来近红外光谱技术在无损检测领域频出贡献^[4],由于其具备分析速度快与样本的非破坏检测等优点,该技术已经成功实现对花椒^[5]、脐橙^[6]、黄酒^[7]、枸杞^[8]、茶叶^[9]、葡萄^[10]、太子参^[11]、枇杷^[12]、雪莲花^[13]等各类产品产地溯源。但是,近红外光谱技术受制于其波段数目较多而存在数据冗余等问题,数据降维的优良将直接影响所建模型的稳定性及成功率。目前常规的数据降维方法主要有主成分分析、独立成分分析、因子分析^[14]等。在此,本文将首次引入在图像识别^[15]、机械故障诊断^[16]等领域取得应用的其他数据降维方法应用于本研究。并且,将高斯过程隐变量模型^[17]、线性局部切空间排列^[18]、近邻元分析^[19]、邻

域保持嵌入^[20]、扩散映射^[21]、t分布随机近邻嵌入^[22]、地标等距映射^[23]、拉普拉斯映射^[24]、局部线性嵌入^[25]在内的多种数据降维方法应用于水果产地溯源模型的建立还未曾见报道。

本文以新疆阿克苏红富士苹果、山东烟台红富士苹果、陕西洛川红富士苹果为研究对象,采用多种预处理方法、数据降维方法和模式识别方法对近红外透射光谱数据分析处理,分别建立基于全波长与特征提取后的分类模型,通过比较不同分类模型的结果可以得到不同光谱预处理方法、不同数据降维方法、不同模式识别方法对苹果产地鉴别的影响。以期对苹果产地溯源的实际应用提供新思路。

1 材料与方法

1.1 苹果样本

样本新疆阿克苏红富士苹果(80°29'E, 41°15'N)于2019年1月6日购买于某水果批发市场;样本山东烟台红富士苹果(121°20'E, 37°33'N)于2019年1月10日购买于某果蔬批发市场、陕西洛川红富士苹果(109°42'E, 35°76'N)于2019年1月10日购买于某果蔬批发市场。由果商对同批次同品牌苹果拆箱挑选大小适中、尺寸均匀、无明显损伤的苹果套网套,打包装箱运回无损检测实验室。光谱采集开始之前,开箱平铺、室温20℃静置24h,擦净苹果表面浮土并逐个编号。将每个产地的200枚苹果样本随机按照1:1的比例分为建模集与预测集,如表1所示。

表1 样本分布

Table 1 Sample Distribution

产地	新疆阿克苏	山东烟台	陕西洛川	总计
建模集	100	100	100	300
预测集	100	100	100	300

1.2 试验设备

近红外透射光谱采集系统如图1。光谱采集范围为590~1250 nm,系统包括苹果托架、配备小型风扇的JCR12V 100 W 卤钨灯光源套件、美国海洋光学公

公司的 USB 2000+型近红外光谱仪、标准 SMA905 接口的光纤、铝合金机架、暗箱与计算机等组成。光纤探头一端连接光谱仪，另一端固定在苹果托架圆心正下方，实现对近红外透射光谱的高效采集。



图 1 近红外透射光谱采集系统

Fig.1 Near Infrared Transmission Spectrum Acquisition System

1.3 试验方法

1.3.1 光谱采集

光谱采集前先将 USB2000+光谱仪开机预热 60 min，使仪器设备达到稳定状态后开始测试采样，通过测试采样设置 SpectraSuite 软件界面参数，得到合适的光谱信息后确定样品光谱采集参数为：平均次数 3；平滑度 5；积分时间 120 ms；波段数 512。采集光谱时，将苹果按照图示置于光谱采集仪器的果托上，注意苹果与果托之间不能留有光缝，确保光纤接收光信号的点完全屏蔽光源，使其只能接收到透过苹果的光。待软件界面显示的光谱稳定后，保存光谱；拿起苹果再经过两次顺时针 120 度旋转，分别采集光谱，最终将三次获得光谱的平均值作为该样本的光谱数据。并且为了消除由于 USB2000+光纤光谱仪预热不充分导致暗光谱发生微小变化所产生的试验误差，所以每测量 10 个样本需保存该时刻的暗光谱用于后续光谱校正。

1.3.2 光谱校正处理

光谱采集过程中，由于苹果形状差异及摄像头中的暗电流的存在会对苹果光谱数据产生噪声影响，需要对获得的近红外透射光谱数据按照公式 1 的方程进行校正：

$$R = \frac{I_{\lambda} - B_{\lambda}}{W_{\lambda} - B_{\lambda}} \quad (1)$$

式中： R 为校正后的光谱； I_{λ} 为原始采集光谱； B_{λ} 为拧上镜头盖采集的全黑暗光谱； W_{λ} 为全反射光谱。

1.4 数据处理方法

1.4.1 预处理方法

采用归一化 (Normalization)、标准正态变量变换 (SNVT)、多元散射校正 (MSC)、卷积平滑法 (SG)、二阶求导 (2nd Derivative)、均值中心化 (MC)、移动平均平滑 (MA) 及一阶求导 (1st Derivative) 对样本的近红外光谱数据做预处理。归一化主要用来消除微小光程差异对光谱变化的影响；标准正态变量变换用来消除样本颗粒大小和光散射等噪声信息夹入；多元散射校正可以缩小由于漫反射造成的光谱数据差异；卷积平滑法可有效消除基线漂移、倾斜；二阶导数可以消弭光谱基线的回旋；均值中心化将光谱的变动而非光谱的绝对量与待测性质或组成的变动进行关联进而提高模型的稳健性；移动窗口平滑可以去除数据变化剧烈的点实现平滑效果；一阶导数可以消除光谱基线的平移。

1.4.2 数据降维方法

采用因子分析 (Factor Analysis)、高斯过程隐变量模型 (GPLVM)、线性局部切空间排列 (LLTSA)、近邻元分析 (NCA)、邻域保持嵌入 (NPE)、扩散映射 (DM)、 t 分布随机近邻嵌入 (t -SNE)、地标等距映射 (L-Isomap)、拉普拉斯映射 (LE)、局部线性嵌入 (LLE)、主成分分析 (PCA)、线性判别分析 (LDA) 方法对样本的预处理后光谱做数据降维处理。

因子分析是从变量中提取共性因子，并要求原有变量间具有较强相关性；高斯过程隐变量模型利用高斯过程对数据进行处理，建立从低维线性流形空间到高维观察数据空间的映射关系，通过最大化观测数据的联合概率密度，优化出高维观察数据在低维空间中的坐标位置；线性局部切空间排列具有自动简化高维混合域特征的能力；近邻元分析作为有监督式数据降维算法，可以最大化区分不同类别间的信息，使同类样本的聚集度更好，不同类样本的区别度更高；邻域保持嵌入在数据降维的同时可以保持样本集原有的局部邻域流形结构；扩散映射通过整合数据的局部几何关系进而揭示数据集在不同尺度的几何结构，其聚焦于发现数据集潜在的流行结构； t 分布随机近邻嵌入在低维空间下，使用 t 分布替代高斯分布表达两点之间的相似度，是在降维的同时考虑数据全局与局部关系的算法；地标等距映射从完整数据中随机选择像元作为标志点，并在每个像元和标志点之间建立最短路径图，其目的在于减少演算量的同时维持标志点与各像元之间的测地间距；拉普拉斯映射通过构建相似关联图来重构数据流形的局部表征结构；局部线性嵌入通过假设局部原始数据近似位于一张超平面上，从而使得该局部的某个数据可以由其邻域数据线性表示；线性判别将高维的样本投影到最佳鉴别矢量空间，以

达到抽取归类信息及压缩特征空间维度的效果。

1.4.3 不同距离度量的K最近邻法

采用不同距离度量的K最近邻法建立产地溯源分类模型。K最近邻法是一种以同类样本在模式空间彼此靠近为依据的算法，它计算在最近邻域中K个已知样本到未知待判样本的距离，然后根据类别，归入比重最大的那一类。而样本在模式空间的相似度由距离度量决定，距离度量方法不同，所对应的模式识别结果也就不同，由于光在苹果组织中的传输特性是一种复杂的非线性结构，所以很难先验确定选用哪种距离度量更为合适^[26]。故此采用4种距离度量方法，分别为欧几里得距离、余弦相似度、城市街区距离、相关距离，计算方法如下：

欧几里得距离：

$$d_{\text{Euclidean}}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

余弦相似度：

$$d_{\text{cosine}}(X, Y) = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2} \sqrt{\sum_{i=1}^n Y_i^2}} \quad (3)$$

城市街区距离：

$$d_{\text{cityblock}}(X, Y) = \sum_{i=1}^n |X_i - Y_i| \quad (4)$$

相关距离：

$$d_{\text{correlation}}(X, Y) = \frac{\text{Cov}(X, Y)}{V_{AR}(X)V_{AR}(Y)} \quad (5)$$

式中：X (X₁, X₂, ..., X_n) 和 Y (Y₁, Y₂, ..., Y_n) 分别为 n 维

空间中的两个样本；C_{OV}为协方差，V_{AR}为方差。

2 结果与讨论

2.1 不同地区苹果平均透射光谱曲线

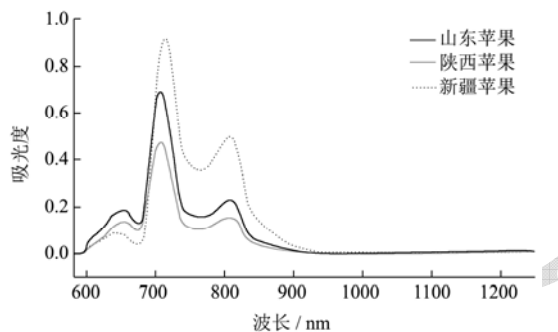


图2 不同地区苹果的近红外透射光谱

Fig.2 Near infrared transmission spectra of apple in different regions

不同地区的苹果物理特性及可溶性固形物含量对比见表 2，其中新疆阿克苏红富士苹果的横径范围及纵径范围最大，陕西洛川苹果最小，平均可溶性固形物含量的次序为新疆>陕西>山东。本文采集了590~1250 nm 波长范围的 512 个波段的近红外数据，每个地区随机挑选 150 个样本的平均光谱经过光谱校正后见图 2。三条平均光谱形状、趋势非常一致，但是新疆苹果在 600~900 nm 与其他两个地区苹果光谱存在差异性分离，而陕西苹果与山东苹果之间的光谱由于吸光度的差异较小而难以区分，因此还需要借助模式识别分类器对以上地区苹果进行分类。

表2 试验苹果常规数据统计

Table 2 Routine Data Statistics of Test Apple

苹果产地	苹果横径/mm	苹果纵径范围/mm	可溶性固形物含量范围/Brix	可溶性固形物平均值/Brix
新疆阿克苏	78.36~102.52	59.44~91.84	12.82~18.26	15.46
山东烟台	73.61~98.12	59.25~90.51	10.67~14.70	12.48
陕西洛川	77.76~94.23	61.59~84.68	10.18~16.72	13.63

2.2 预处理方法的选取

原始光谱数据所获取的信息除了有含有待测苹果的原始信息之外，还包含各种外在干扰信息，如检测环境、样本背景以及设备自身的干扰。为尽可能减小以上误差对数据分析产生的干扰，同时也为筛选出最佳预处理方法，采用 8 种预处理方法对光谱数据进行预处理之后，再使用基于欧几里得距离作为距离度量的K最近邻法对经以上预处理方法后的光谱数据建模

分析，结果如表 3。

由表 3 可见，基于不同预处理方法的 K 最近邻法模型建模正确率区间为 84.60%~89.60%，预测集正确率区间为 75.60%~83.60%。可见利用近红外透射光谱技术对苹果产地进行鉴别的方案可行。综合考虑可知最优预处理方法为二阶求导，最优的 K 取值为 7。但是由于全光谱建模耗时长、变量复杂并且不稳定等缺点，所以并不能直接应用于生产实际中，故选取二阶求导预处理对样本集合做进一步分析。

表3 全波长下基于不同预处理方法的KNN模型苹果产地识别率

Table 3 Apple origin recognition rate of KNN model based on different pretreatment methods under full spectra

预处理方法	K 取值	建模集精度/%	预测集精度/%
归一化	7	84.60	77.00
标准正态变量变换	5	84.60	75.60
多元散射校正	5	86.30	76.30
卷积平滑法	7	85.00	77.30
二阶求导	7	89.60	83.60
均值中心化	7	85.30	77.30
移动窗口平滑	7	84.60	76.60
一阶求导	7	89.60	81.30
原始	7	85.30	77.30

2.3 基于全波段的分类识别结果

表4 基于全波段的四种模型苹果产地识别率

Table 4 Apple origin recognition rate based on four models under full spectra

模型	建模集精度/%	预测集精度/%
K 最近邻法(Euclidean)	89.60	83.60
K 最近邻法(cosine)	90.30	81.30
K 最近邻法(cityblock)	91.30	84.30
K 最近邻法(correlation)	92.00	85.60

对光谱数据做二阶求导预处理，并且基于不同度量距离的 K 最近邻法进行全波长建模分析，旨在找出最优模式识别方法，识别结果如表 4。以欧几里得距离为距离度量的 K 最近邻法预测精度是最低的，分类效果最好的是以相关距离为距离度量的 K 最近邻法，建模集正确率为 92.0%，预测集正确率为 85.6%。

表5 基于不同数据降维方法的K最近邻法模型苹果产地识别率

Table 5 Apple origin recognition rate With KNN model based on different data dimension reduction methods

降维方法	主成分数	建模集精度/%	预测集精度/%
因子分析	11	89.30	87.00
高斯过程隐变量模型	8	92.60	88.00
线性局部切空间排列	4	86.30	83.30
近邻元分析	13	92.30	86.00
邻域保持嵌入	7	90.30	83.00
扩散映射	9	97.30	92.30
t 分布随机近邻嵌入	13	94.00	84.60
地标等距映射	10	87.30	82.30
拉普拉斯映射	5	86.60	82.60
局部线性嵌入	9	85.30	81.00
主成分分析	13	94.30	89.30
线性判别分析	13	85.60	82.30

2.4 基于多种数据降维方法下的模式识别分类结果

本文采集的光谱数据包含 512 个波段，若不对其进行数据降维，模型的识别将会耗费大量的时间并且不能保证是否有无关信息对预测精度造成不良影响。因此，为了对比不同数据降维方法对近红外透射光谱下苹果产地溯源模型的影响；本文将经过二阶求导预处理后的光谱数据结合多种数据降维方法对光谱数据做降维处理，再将所有降到合适维度的数据结合前文所选最优分类方法，即 K 最近邻法 (correlation) 进行建模分析，在建模分析中 1 被用来代表代表新疆红富士苹果, 2 为陕西红富士苹果, 3 为山东红富士苹果。所得结果如表 5。

由表 5 可知,在二阶求导预处理下,数据降维效果最好的是扩散映射,提取前 9 个主成分,建模集和预测集的分类识别率分别为 97.30%与 92.30%。经过数据降维之后的光谱数据,结合 K 最近邻法 (correlation) 建模不仅在光谱维度上进行了合理约减,在识别正确率方面也要比全波长下建模效果好。

不同产地的红富士苹果在近红外光谱上存在的差异主要由于不同产地红富士苹果自身的有机物的组成及含量不同;其次,不同产地的红富士苹果由于温湿度、海拔高度、日照时间、平均温度、年降水量等众多外界因素影响也会造成一定的差异。本试验利用近红外光谱技术结合深度学习算法使得利用较少的数据信息便可得到较优结果,过程更加快捷、准确。经过预处理后的光谱数据结合欧几里得距离、相关距离、余弦相似度、城市街区距离作为距离度量的 K 最近邻法分类最优正确率达到 92.30%,证明近红外透射光谱技术结合化学计量算法对苹果产地分鉴具有可行性。

3 结论

3.1 提出一种基于深度学习数据降维方法的苹果产地鉴别方法。对采集后经光谱校正处理后的光谱数据用归一化、标准正态变量变换、多元散射校正、卷积平滑法、二阶求导、均值中心化、移动窗口平滑及一阶求导进行光谱预处理;随后对所有光谱预处理后的光谱数据建立全波长模式识别模型,结果表明二阶求导预处理方法最好,所建模型的建模集和预测集的分类识别率分别为 89.60%与 83.60%。

3.2 对经过二阶求导预处理方法之后的光谱数据结合欧几里得距离、相关距离、余弦相似度、城市街区距离作为距离度量的 K 最近邻法进行苹果产地溯源的建模分析;发现以相关距离作为距离度量的 K 最近邻法为最优分类识别方法,所建模型的建模集和预测集的分类识别率分别为 92.00%与 85.60%,其中最优 K 取值为 7。

3.3 结合因子分析、高斯过程隐变量模型、线性局部切空间排列、近邻元分析、邻域保持嵌入、扩散映射、t 分布随机近邻嵌入、地标等距映射、拉普拉斯映射、局部线性嵌入、主成分分析、线性判别分析算法对经过二阶求导预处理后的数据集做降维处理,再将降维处理后的光谱数据结合 K 最近邻法 (correlation) 建模。发现扩散映射为最优降维方法,其能在保持数据内在几何结构的同时对其进行非线性降维,对数据有聚类效果。扩散映射提取前 9 个主成分,数据建模集和预测集的分类识别率分别为 97.30%与 92.30%。

3.4 本方法为红富士苹果产地溯源提供技术支持,可

以有效控制各产地苹果乱掺乱卖、以次充好的卑劣手段,为保护具有国家地理标志的苹果产品提供检测识别的新思路。

参考文献

- [1] 中国科学院中国植物志编辑委员会.中国植物志[M].北京:科学出版社,2004:371-372
Editorial Board of Chinese Flora of Chinese Academy of Sciences. Flora of China [M]. Beijing: Science Press, 2004: 371-372
- [2] 孙平平,王文辉.2017/2018 年世界苹果、梨、葡萄、桃及樱桃产量、市场与贸易情况[J].中国果树,2018,2:99-108
SUN Ping-ping, WANG Wen-hui. World apple, pear, grape, peach and cherry production, market and trade in 2017/2018 [J]. China Fruits, 2018, 2: 99-108
- [3] 冯娟,任小林,田建文,等.不同产地富士苹果品质分析与比较[J].食品工业科技,2013,34(14):108-112
FENG Juan, REN Xiao-lin, TIAN Jian-wen, et al. Analysis and comparison of Fuji apple quality from different regions [J]. Science and Technology of Food Industry, 2013, 34(14): 108-112
- [4] 邹昊,田寒友,刘飞,等.近红外光谱的预处理对羊肉 TVB-N 模型的影响[J].食品科学,2016,37(22):180-186
ZOU Hao, TIAN Han-you, LIU Fei, et al. Effects of spectral pretreatments on prediction of total volatile basic nitrogen (TVB-N) content in mutton using near infrared spectroscopy [J]. Food Science, 2016, 37(22): 180-186
- [5] 吴习宇,祝诗平,黄华,等.近红外光谱技术鉴别花椒产地[J].光谱学与光谱分析,2018,38(1):68-72
WU Xi-yu, ZHU Shi-ping, HUANG Hua, et al. Near infrared spectroscopy for determination of the geographical origin of Huajiao [J]. Spectroscopy and Spectral Analysis, 2018, 38(1): 68-72
- [6] 苏学素,张晓焱,焦必宁,等.基于近红外光谱的脐橙产地溯源研究[J].农业工程学报,2012,28(15):240-245
SU Xue-su, ZHANG Xiao-yan, JIAO Bi-ning, et al. Determination of geographical origin of navel orange by near infrared spectroscopy [J]. Modern Food Science and Technology, 2012, 28(15): 240-245
- [7] Shen F, Yang D, Ying Y, et al. Discrimination between shaoxing wines and other chinese rice wines by near-infrared spectroscopy and chemometrics [J]. Food and Bioprocess Technology, 2012, 5(2): 786-795
- [8] 李仲,刘明地,吉守祥.基于枸杞红外光谱人工神经网络的产地鉴别[J].光谱学与光谱分析,2016,36(3):720-723

- LI Zhong, LIU Ming-de, JI Shou-xiang. The identification of the origin of Chinese wolfberry based on infrared spectral technology and the artificial neural network [J]. *Spectroscopy and Spectral Analysis*, 2016, 36(3): 720-723
- [9] Chen Q, Zhao J, Lin H. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition [J]. *Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy*, 2009, 72(4): 845-850
- [10] Arana I, Arazuri S. Maturity, variety and origin determination in white grapes (*Vitis vinifera* L.) using near infrared reflectance technology [J]. *Journal of Near Infrared Spectroscopy*, 2005, 13(1): 349
- [11] Lin H, Zhao J, Chen Q, et al. Discrimination of radix pseudostellariae according to geographical origins using NIR spectroscopy and support vector data description [J]. *Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy*, 2011, 79(5): 1381-1385
- [12] Fu X, Ying Y, Zhou Y, et al. Application of probabilistic neural networks in qualitative analysis of near infrared spectra: Determination of producing area and variety of loquats [J]. *Analytica Chimica Acta*, 2007, 598(1): 27-33
- [13] 赵杰文, 蒋培, 陈全胜. 雪莲花产地鉴别的近红外光谱分析方法[J]. *农业机械学报*, 2010, 41(8): 111-114
ZHAO Jie-wen, JIANG Pei, CHEN Quan-sheng. Discrimination of snow lotus from different geographical origins by near infrared spectroscopy [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2010, 41(8): 111-114
- [14] Li C, Li E, Zhang C, et al. Factor analysis based quantitative determination of flavor type and its corresponding relationship with growing areas in flue-cured tobacco [J]. *Acta Tabacaria Sinica*, 2016, 22(6): 51-62
- [15] Wang F, Zhang H, Zuo W, et al. Face recognition using fast neighborhood component analysis with spatially smooth regularizer [J]. *Lecture Notes in Computer Science*, 2012, 7751: 450-457
- [16] 陈法法, 杨晓青, 陈保家, 等. 基于正交邻域保持嵌入与多核相关向量机的滚动轴承早期故障诊断[J]. *计算机集成制造系统*, 2018, 24(8): 1946-1954
CHEN Fa-fa, YANG Xiao-qing, CHEN Bao-jia, et al. Early fault diagnosis of rolling bearing based on orthogonal neighborhood preserving embedding and multi-kernel relevance vector machine [J]. *Computer Integrated Manufacturing Systems*, 2018, 24(8): 1946-1954
- [17] Han B, Zhang L, Gao X, et al. Embedded locality discriminant GPLVM for dimensionality reduction [C]. *International Joint Conference on Neural Networks*. IEEE, 2016
- [18] 李锋, 汤宝平, 陈法法. 基于线性局部切空间排列维数化简的故障诊断[J]. *振动与冲击*, 2012, 31(13): 36-40, 61
LI Feng, TANG Bao-ping, CHEN Fa-fa. Fault diagnosis model based on dimension reduction using linear local tangent space alignment [J]. *Journal of Vibration and Shock*, 2012, 31(13): 36-40, 61
- [19] Liu X, Wan H, Shi L. Quality metrics of spike sorting using neighborhood components analysis [J]. *The Open Biomedical Engineering Journal*, 2014, 8(1): 60-67
- [20] Yuan X, Ge Z, Ye L, et al. Supervised neighborhood preserving embedding for feature extraction and its application for soft sensor modeling: Supervised neighborhood preserving embedding [J]. *Journal of Chemometrics*, 2016, 30(8): 430-441
- [21] 倪家鹏, 沈韬, 朱艳, 等. 基于扩散映射的太赫兹光谱识别[J]. *光谱学与光谱分析*, 2017, 37(8): 2360-2364
NI Jia-peng, SHEN Tao, ZHU Yan, et al. Terahertz spectroscopic identification with diffusion maps [J]. *Spectroscopy and Spectral Analysis*, 2017, 37(8): 2360-2364
- [22] 徐森, 花小朋, 徐静, 等. 一种基于 T-分布随机近邻嵌入的聚类集成方法[J]. *电子与信息学报*, 2018, 40(6): 1316-1322
XU Sen, HUA Xiao-peng, XU Jing, et al. Cluster ensemble approach based on T-distributed stochastic neighbor embedding [J]. *Journal of Electronics & Information Technology*, 2018, 40(6): 1316-1322
- [23] Orsenigo, Carlotta. An improved set covering problem for isomap supervised landmark selection [J]. *Pattern Recognition Letters*, 2014, 49: 131-137
- [24] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. *Neural Computation*, 2003, 15(6): 1373-1396
- [25] Liu X, Tosun D, Weiner M W, et al. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification [J]. *Neuro Image*, 2013, 83: 148-157
- [26] 马惠玲, 王若琳, 蔡骋, 等. 基于高光谱成像的苹果品种快速鉴别[J]. *农业机械学报*, 2017, 48(4): 305-312
MA Hui-ling, WANG Ruo-lin, CAI Cheng, et al. Rapid identification of apple varieties based on hyperspectral imaging [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(4): 305-312